

Lecture Notes in Bioinformatics

7986

Edited by S. Istrail, P. Pevzner, and M. Waterman

Editorial Board: A. Apostolico S. Brunak M. Gelfand

T. Lengauer S. Miyano G. Myers M.-F. Sagot D. Sankoff

R. Shamir T. Speed M. Vingron W. Wong

Subseries of Lecture Notes in Computer Science

Alioune Ngom Enrico Formenti
Jin-Kao Hao Xing-Ming Zhao
Twan van Laarhoven (Eds.)

Pattern Recognition in Bioinformatics

8th IAPR International Conference, PRIB 2013
Nice, France, June 17-20, 2013
Proceedings



Springer

Volume Editors

Alioune Ngom
University of Windsor, ON, Canada
E-mail: angom@uwindsor.ca

Enrico Formenti
Nice Sophia Antipolis University, France
E-mail: enrico.formenti@unice.fr

Jin-Kao Hao
Université d'Angers, France
E-mail: hao@info.univ-angers.fr

Xing-Ming Zhao
Tongji University, Shanghai, China
E-mail: zhaoxingming@gmail.com

Twan van Laarhoven
Radboud University, Nijmegen, The Netherlands
E-mail: tvanlaarhoven@cs.ru.nl

ISSN 0302-9743	e-ISSN 1611-3349
ISBN 978-3-642-39158-3	e-ISBN 978-3-642-39159-0
DOI 10.1007/978-3-642-39159-0	
Springer Heidelberg Dordrecht London New York	

Library of Congress Control Number: 2013941044

CR Subject Classification (1998): J.3, I.5, F.2.2, I.2, I.4, H.3.3, H.2.8

LNCS Sublibrary: SL 8 – Bioinformatics

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

In the post-genomic era, a holistic understanding of biological systems and processes, in all their complexity, is critical in comprehending nature's choreography of life. As a result, bioinformatics involving its two main disciplines, namely, the life sciences and the computational sciences, is fast becoming a very promising multidisciplinary research field. With the ever-increasing application of large-scale high-throughput technologies, such as gene or protein microarrays and mass spectrometry methods, the enormous body of information is growing rapidly. Bioinformaticians are posed with a large number of difficult problems to solve, arising not only due to the complexities in acquiring the molecular information but also due to the size and nature of the generated data sets and/or the limitations of the algorithms required for analyzing these data. The recent advancements in computational and information-theoretic techniques are enabling us to conduct various *in silico* testing and screening of many lab-based experiments before these are actually performed *in vitro* or *in vivo*. These *in silico* investigations are providing new insights for interpreting and establishing new direction for a deeper understanding. Among the various advanced computational methods currently being applied to such studies, the *pattern recognition* techniques are mostly found to be at the core of the whole discovery process for apprehending the underlying biological knowledge. Thus, we can safely surmise that the ongoing bioinformatics *revolution* may, in future, inevitably play a major role in many aspects of medical practice and/or the discipline of life sciences.

The aim of this conference on Pattern Recognition in Bioinformatics (PRIB) is to provide an opportunity to academics, researchers, scientists, and industry professionals to present their latest research in pattern recognition and computational intelligence-based techniques applied to problems in bioinformatics and computational biology. It also provides them with an excellent forum to interact with each other and share experiences. The conference is organized jointly by the Nice Sophia Antipolis University, France, and IAPR (International Association for Pattern Recognition) Bioinformatics Technical Committee (TC-20).

This volume presents the proceedings of the 8th IAPR International Conference on Pattern Recognition in Bioinformatics (PRIB 2013), held in Nice, June 17–19, 2013. It includes 25 technical contributions that were selected by the International Program Committee from 43 submissions. Each of these rigorously reviewed papers was presented orally at PRIB 2013. The proceedings consists of five parts:

Part I Bio-Molecular Networks and Pathway Analysis

Part II Learning, Classification, and Clustering

Part III Data Mining and Knowledge Discovery

Part IV Protein: Structure, Function, and Interaction

Part V Motifs, Sites, and Sequences Analysis

Part I of the proceedings contains six chapters on “Bio-Molecular Networks and Pathway Analysis.” Rahman et al. propose a fast agglomerative clustering method for protein complex discovery. A new criterion is introduced that combines an edge clustering coefficient and an edge clustering value, allowing us to decide when a node can be added to the current cluster. Maduranga et al. use the well-known random forest method to predict GRNs. The problem of inferring GRNs from (limited) time-series data is recast as a number of regression problems, and the random forest approach is used here to fit a model to this. Winterbach et al. evaluate how well topological signatures in protein interaction networks predict protein function. They compare several complex signatures and their own simple signature. They find that network topology is only a weak predictor of function and the simple signature performs on par with the more sophisticated ones. De Ridder et al. propose an approach for identifying putative cancer pathways. This approach relies on expression profiling tumors that are induced by retroviral insertional mutagenesis. This provides the opportunity to search for associations between tumor-initiating events (the viral insertion sites) and the consequent transcription changes, thus revealing putative regulatory interactions. An important advantage is that the selective pressure exerted by the tumor growth is exploited to yield a relatively small number of loci that are likely to be causal for tumor formation. Ochs et al. apply outlier statistics, gene set analysis, and top scoring pair methods to identify deregulated pathways in cancer. Analysis of the results on pediatric acute myeloid leukemia data indicate the effectiveness of the proposed methodology. Pizzuti et al. present some variants of RNSC (restricted neighborhood search clustering) for prediction of protein complexes that are based on new score functions and evolutionary computation. It is shown via computational experiments that the proposed methods have better prediction accuracies (in F-measure) than the basic RNSC algorithm.

Part II of the proceedings contains three chapters on “Learning, Classification, and Clustering.” Marchiori addresses a limitation of the RELIEF feature weighting algorithm that maximizes the sample margin over the entire training set, or the sum of the possibly competing feature weights. Her work proposes, instead, a conditional weighting algorithm (CCFW) and classifier (CCWNN) to improve feature weighting and classification. Mundra et al. propose a sample selection criterion using a modified logistic regression loss function and a backward elimination based gene ranking algorithm. On the basis of the classifier margin for sample points, points on or within the margin are more important than those outside, the sample selection criterion based on T-score is proposed. Li et al. describe a generalization of sparse matrix factorization (SMF) algorithms and showcase a few very concisely described applications in bioinformatics. The main merit of the work is the fact that a unified representation for SMF algorithms is proposed, as well as an optimization algorithm to solve this problem.

Part III of the proceedings contains six chapters on “Data Mining and Knowledge Discovery.” Hsu et al. consider prediction of RNA secondary structure in the “triple helix” setting for which they argue existing methods are inadequate. Their approach uses a Simple Tree Adjoining Grammar (STAG) coupled

with maximum likelihood estimation (MLE), implemented via an efficient dynamic programming formulation. Higgs et al. present an algorithm for generating near-native protein models. It combines a fragment feature-based resampling algorithm with a local optimization method that performed best, for protein structure prediction (PSP), among a set of five optimization techniques. Computational experiments show that the use of local optimization is beneficial in terms of both RMSD and TM score. Spirov et al. discuss a method for transformation of variables, in order to normalize *Drosophila oocyte* images acquired via confocal microscopy. The paper describes an interesting problem, namely, the experimental determination of intrinsic *Drosophila* embryo coordinates, and proposes an approach using evolutionary computation by genetic algorithms. Rezaeian et al. propose a novel and flexible hierarchical framework to select discriminative genes and predict breast tumor subtypes simultaneously. Dai et al. tackle an important problem in drug-target interaction research and present an interesting application of machine learning methods to the analysis of drugs. Gritsenko et al. make an adaptation of their previously developed protocol for building and evaluating predictors, in order to introduce a framework that enables forward engineering in biology. An experimental test is performed in the biological field of codon optimization and the results obtained are comparable with those produced by the reference tool JCat.

Part IV of the proceedings contains six chapters on “Protein: Structure, Function, and Interaction.” Xiong et al. propose an active learning-based approach for protein function prediction. The novelty of the proposal is the use of a pre-processing phase that uses spectral clustering before selecting candidates for labeling with graph centrality metrics. Experimental results show that clustering reveals a valid pre-processing step for the active learning method. Gehrmann et al. address the problem of integrating multiple sources of evidence to predict protein functions. The paper proposes to use a conditional random field (CRF) to represent protein functions as random variables to be predicted and different sources of evidence as conditioning variables. Inference and learning algorithms based on MCMC are described and the proposed method is applied to a yeast dataset. Dehzangi et al. describe a new approach to protein fold recognition, a problem that has been widely studied over the past decade. The main contribution is the proposal of a new set of global protein features based on evolutionary consensus sequences and predicted secondary structure, and local features based on distributions and auto covariances of these features over segments. An RBF SVM using these features is applied to two benchmark datasets in an extensive comparison with a number of existing methods and is demonstrated to work well. Dehzangi et al. present a novel approach to using features extracted from the position specific scoring matrix (PSSM) to predict the structural class of a protein. The authors propose two new sets of features: a global one based on the consensus sequence of a PSSM and a local one that takes the auto-covariance in sequence segments into account. The features extracted are used to train an RBF SVM and are shown to lead to good results (better than other state-of-the-art algorithms) on two benchmarks. Chiu et al. discuss a new method for detecting

associated sites in aligned sequence ensembles. The main idea is derived from the concept of granular computing, where information is extracted at different levels of granularity or resolution. The experimentation was focused on p53 and it has been demonstrated that the extracted association patterns are useful in discovering sites with some structural and functional properties of a protein molecule. Tung presents a new method for predicting the potential hepatocarcinogenicity of non-genotoxic chemicals. The proposed method based on chemical-protein interactions and interpretable decision tree is compared with other data-mining approaches and shows very good performances in both accuracy and simplicity of the found model.

Part V of the proceedings contains four chapters on “Motifs, Sites, and Sequences Analysis.” Pathak et al. present an algorithm that exploits structural information for reducing false positives in motifs prediction. They tested the validity of the algorithm using the minimotifs stored in the MnM database. Lacroix et al. present a workflow for the prediction of the effects of residue substitution on protein stability. The workflow integrates eight algorithms that use delta-delta-G as a measure of stability. The workflow is designed to populate the online resource SPROUTS. A use case of the workflow is presented using the PDB entry 1enh. Malhotra et al. present an algorithm for inferring haplotypes of virus populations from k-mer counts obtained from next-generation sequencing (NGS) data. The algorithm takes as input read counts for a set of k-mers and produces as output a predicted number of haplotypes, their relative frequencies and, for reads covering SNPs, can assign reads to a haplotype. The novel feature of the algorithm is that it does not rely on having a reference genome. The authors report that it performs well on synthetic data compared with the existing algorithm ShoRAH, which relies on a reference genome. Comin et al. discuss and improve the Entropic Profile method introduced in the literature for detecting conservation in genome sequences. The authors propose a linear-time linear-space algorithm that captures the importance of given regions with respect to the whole genome, suitable for large genomes and for the discovery of motifs with unbounded length.

Many have contributed directly or indirectly toward the organization and success of the PRIB 2013 conference. We would like to thank all the individuals and institutions, especially the authors for submitting the papers and the sponsors for generously providing financial support for the conference. We are very grateful to IAPR for the sponsorship. Our gratitude goes to the Nice Sophia Antipolis University, Nice, France, and IAPR (International Association for Pattern Recognition) Bioinformatics Technical Committee (TC-20) for supporting the conference in many ways.

We would like to express our gratitude to all PRIB 2013 International Program Committee members for their objective and thorough reviews of the submitted papers. We fully appreciate the PRIB 2013 Organizing Committee for their time, efforts, and excellent work. We would also like to thank the Nice Sophia Antipolis University for hosting the symposium and providing technical support. We sincerely thank the EDSTIC doctoral school for providing grants to

a number of students attending the conference. We also thank “Region PACA” and the University of Salerno (Italy) for partially funding the invited speakers.

Last, but not least, we wish to convey our sincere thanks to Springer for providing excellent professional support in preparing this volume.

June 2013

Alioune Ngom
Enrico Formenti
Jin-Kao Hao
Xing-Ming Zhao
Twan van Laarhoven

PRIB 2013 Organization

PRIB General Chair

Alioune Ngom

University of Windsor, Canada

CIBB-PRIB General Chair

Enrico Formenti

Nice Sophia Antipolis University, France

Program Co-chairs

Jin-Kao Hao

University of Angers, France

Xing-Ming Zhao

Tongji University, China

Publicity Chair

Gilles Bernot

Nice Sophia Antipolis University, France

Local Arrangements Chair

Jean-Paul Comet

Nice Sophia Antipolis University, France

Publication Chair

Twan van Laarhoven

Radboud University, The Netherlands

PRIB Steering Committee

Raj Acharya

PennState, USA

Shandar Ahmad

Tokyo Tech NIBIO, Japan

Madhu Chetty

Monash University, Australia

Tom Heskes

Radboud University of Nijmegen,

The Netherlands

Visakan Kadiramanathan

University of Sheffield, UK

Elena Marchiori

Radboud University of Nijmegen,

The Netherlands

Jagath C. Rajapakse

Nanyang Technological University, Singapore

Marcel Reinders

Delft University of Technology,

The Netherlands

Dick de Ridder

Delft University of Technology,

The Netherlands

Guido Sanguinetti

University of Edinburgh, UK

Jun Sese

Tokyo Tech NIBIO, Japan

Program Committee

Raj Acharya	Pennsylvania State University, USA
Shandar Ahmad	National Institute of Biomedical Innovation, Osaka
Tatsuya Akutsu	Kyoto University, Japan
Sahar Al Seesi	University of Connecticut, USA
Hisham Al-Mubaid	University of Houston, USA
Ashish Anand	Institut Pasteur Paris, France
Kiyoko Aoki-Kinoshita	Soka University, Japan
Wendy Ashlock	University of Guelph, Canada
Francisco Azuaje	Public Research Centre for Health, UK
Jaume Bacardit	University of Nottingham, UK
Pedro Ballester	Cambridge University, UK
Sanghamitra Bandyopadhyay	Indian Statistical Institute, India
Gilles Bernot	Nice Sophia Antipolis University, France
Chengpeng Bi	University of Missouri, UK
Sebastian Böcker	Friedrich Schiller University of Jena, Germany
Conrad Burden	Australian National University, Australia
David Cairns	University of Stirling, UK
Rachel Cavill	Imperial College London, UK
Frederic Cazals	INRIA Sophia, France
Keith C.C. Chan	The Hong Kong Polytechnic University, China
Kuo-Sheng Cheng	National Cheng Kung University, Taiwan
Francis Chin	The University of Hong Kong, China
Sung-Bae Cho	Yonsei University, Korea
Young-Rae Cho	Baylor University, USA
Dominique Chu	University of Kent, UK
Pau-Choo Chung	National Cheng Kung University, Taiwan
Steven Corns	Missouri University of Science and Technology, USA
Sanjoy Das	Kansas State University, USA
Dick De Ridder	Delft University of Technology, The Netherlands
Jeroen De Ridder	Delft University of Technology, The Netherlands
Tjeerd Dijkstra	Radboud University, The Netherlands
Federico Divina	Pablo de Olavide University, Spain
Beatrice Duval	University of Angers, France
Mansour Ebrahimi	University of Qom, Iran
Esmail Ebrahimie	Shiraz University, Iran
Richard Edwards	University of Southampton, UK
Antonino Fiannaca	ICAR-CNR, Italy
Maurizio Filippone	University of Glasgow, UK

Christoph M. Friedrich	University of Applied Science and Arts, Germany
Rosalba Giugno	University of Catania, Italy
Robin Gras	University of Windsor, Canada
Michael Gromiha	IIT Madras, India
Michael Hahsler	Southern Methodist University, USA
Jennifer Hallinan	Newcastle University, UK
Xiaoxu Han	University of Iowa, USA
Timothy Havens	University of Missouri, USA
Morihiro Hayashida	Kyoto University, Japan
David Hecht	Southwestern College, USA
Md Tamjidul Hoque	Griffith University, Australia
Sheridan Houghten	Brock University, Canada
Liang-Tsung Huang	Mingdao University, Taiwan
Seiya Imoto	University of Tokyo, Japan
Zhenyu Jia	University of California Irvine, USA
Colin Johnson	University of Kent, UK
Laetitia Jourdan	INRIA, France
David Juedes	Ohio University, USA
Giuseppe Jurman	Fondazione Bruno Kessler, Italy
R. Krishna Murthy Karuturi	Genome Institute of Singapore, Singapore
Marta Kasprzak	Poznan University of Technology, Poland
Yuki Kato	Nara Institute of Science and Technology, Japan
Tsuyoshi Kato	University of Tokyo, Japan
Nawaz Khan	Middlesex University, UK
Seyoung Kim	Carnegie Mellon University, USA
Kyung Dae Ko	Howard University, USA
Ziad Kobti	University of Windsor, Canada
Tetsuji Kuboyama	Gakushuin University, Japan
Lukasz Kurgan	University of Alberta, Canada
Zoe Lacroix	Arizona State University, USA
Yifeng Li	University of Windsor, Canada
Xiaoli Li	Institute for Infocomm Research, Singapore
Wingning Li	University of Arkansas, USA
Feng Lin	Nanyang Technological University, Singapore
Frédérique Lisacek	Swiss Institute of Bioinformatics, Switzerland
Chunmei Liu	Howard University, USA
Xuejun Liu	Nanjing University of Aeronautics and Astronautics, China
Weiguo Liu	Nanyang Technological University, Singapore
Huaien Luo	Genome Institute of Singapore, Singapore
Hiroshi Matsuno	Yamaguchi University, Japan
Ken McGarry	University of Sunderland, UK

Vasilis Megalooikonomou	Temple University, USA
Mariofanna Milanova	University of Arkansas at Little Rock, USA
Aleksandar Milosavljevic	Baylor College of Medicine, USA
Perry Moerland	Academic Medical Center, The Netherlands
Jason Moore	Dartmouth College, USA
Vadim Mottl	Computing Center of the Russian Academy of Sciences, Russia
Piyushkumar Mundra	Nanyang Technological University, Singapore
Julio Cesar Nievola	Pontifícia Universidade Católica do Paraná, Brazil
Michael Ochs	Johns Hopkins University, USA
Carlotta Orsenigo	Politecnico di Milano, Italy
Deanna Petrochilos	University of Washington, USA
Thang Pham	VU University Medical Center, The Netherlands
Clara Pizzuti	CNR-ICAR, Italy
Kiran Sree Pokkuluri	Venkata Chalamayya Engineering College, India
Gianfranco Mihele	
Maria Politano	Politecnico di Torino, Italy
Beatriz Pontes	University of Seville, Spain
Sanguthevar Rajasekaran	University of Connecticut, USA
Jean-Michel Richer	University of Angers, France
Ricardo Rizzo	National Research Council of Italy, Italy
Miguel Rocha	University of Minho, Portugal
Katya Rodriguez-Vazquez	IIMAS-UNAM, Mexico
Eduardo Rodriguez-Tello	Cinvestav-Tamaulipas, Mexico
Luis Rueda	University of Windsor, Canada
Gonzalo Ruz	Universidad Adolfo Ibáñez, Chile
Yvan Saeys	Ghent University, Belgium
Hiroto Saigo	Kyushu Institute of Technology, Japan
Taro L. Saito	University of Tokyo, Japan
Roberto Santana	Technical University of Madrid, Spain
Bertil Schmidt	University of Mainz, Germany
Christian Schönbach	Kyushu Institute of Technology, Japan
Huseyin Seker	De Montfort University, UK
Masakazu Sekijima	Tokyo Institute of Technology, Japan
Jun Sese	Tokyo Institute of Technology, Japan
Anne Siegel	IRISA – CNRS, France
Luciano Silva	Universidade Federal do Parana, Brazil
Evangelos Simeonidis	Luxembourg Centre for Systems Biomedicine, Luxembourg
Alexander Spirov	SUNY at Stony Brook, USA
Bela Stantic	Griffith University, Australia
Gregor Stiglic	University of Maribor, Slovenia

Chrysostomos Stylios	Technological Education Institute of Epirus, Greece
Wing-Kin Sung	Nuational University of Singapore, Singapore
Kenji Suzuki	The University of Chicago, USA
Marta Szachniuk	Polish Academy of Sciences, Poland
Sandor Szilagyi	Sapientia-Hungarian Science University of Transylvania, Hungary
Roberto Tagliaferri	University of Salerno, Italy
Shyh Wei Teng	Monash University, Australia
Spencer Thomas	University of Surrey, UK
Renato Tinós	University of Sao Paulo, Brazil
Anna Tramontano	University of Rome La Sapienza, Italy
Herbert Treutlein	Computist Bio-Nanotech, Australia
Herbert H. Tsang	Simon Fraser University, Canada
Alexey Tsymbal	Siemens AG, Germany
Chun-Wei Tung	National Chiao Tung University, Taiwan
Marcel Turcotte	University of Ottawa, Canada
Alfonso Urso	Consiglio Nazionale delle Ricerche, Italy
Jean-Philippe Vert	Ecole des Mines de Paris, France
Jiri Vohradsky	Institute of Microbiology ASCR, Czech Republic
Junbai Wang	Radium Hospital, Norway
Lusheng Wang	City University of Hong Kong, Hong Kong
Haixuan Yang	Royal Holloway University of London, UK
Junichiro Yoshimoto	Okinawa Institute of Science and Technology, Japan
Yanqing Zhang	Georgia State University, USA
Daming Zhu	Shandong University, China

Table of Contents

Part I: Bio-molecular Networks and Pathway Analysis

A Fast Agglomerative Community Detection Method for Protein Complex Discovery in Protein Interaction Networks	1
<i>Mohammad S. Rahman and Alioune Ngom</i>	
Inferring Gene Regulatory Networks from Time-Series Expressions Using Random Forests Ensemble	13
<i>D.A.K. Maduranga, Jie Zheng, Piyushkumar A. Mundra, and Jagath C. Rajapakse</i>	
Local Topological Signatures for Network-Based Prediction of Biological Function	23
<i>Wynand Winterbach, Piet Van Mieghem, Marcel J.T. Reinders, Huijuan Wang, and Dick de Ridder</i>	
Mutational Genomics for Cancer Pathway Discovery	35
<i>Jeroen de Ridder, Jaap Kool, Anthony G. Uren, Jan Bot, Johann de Jong, Alistair G. Rust, Anton Berns, Maarten van Lohuizen, David J. Adams, Lodewyk Wessels, and Marcel J.T. Reinders</i>	
Outlier Gene Set Analysis Combined with Top Scoring Pair Provides Robust Biomarkers of Pathway Activity	47
<i>Michael F. Ochs, Jason E. Farrar, Michael Considine, Yingying Wei, Soheil Meschinchì, and Robert J. Arceci</i>	
Restricted Neighborhood Search Clustering Revisited: An Evolutionary Computation Perspective	59
<i>Clara Pizzuti and Simona E. Rombo</i>	

Part II: Learning, Classification, and Clustering

Class Dependent Feature Weighting and K-Nearest Neighbor Classification	69
<i>Elena Marchiori</i>	
Simultaneous Sample and Gene Selection Using T-score and Approximate Support Vectors	79
<i>Piyushkumar A. Mundra, Jagath C. Rajapakse, and D.A.K. Maduranga</i>	

Versatile Sparse Matrix Factorization and Its Applications in High-Dimensional Biological Data Analysis	91
<i>Yifeng Li and Alioune Ngom</i>	

Part III: Data Mining and Knowledge Discovery

A Local Structural Prediction Algorithm for RNA Triple Helix Structure	102
<i>Bay-Yuan Hsu, Thomas K.F. Wong, Wing-Kai Hon, Xinyi Liu, Tak-Wah Lam, and Siu-Ming Yiu</i>	
Combining Protein Fragment Feature-Based Resampling and Local Optimisation	114
<i>Trent Higgs, Lukas Folkman, and Bela Stantic</i>	
Experimental Determination of Intrinsic Drosophila Embryo Coordinates by Evolutionary Computation	126
<i>Alexander V. Spirov, Carlos E. Vanario-Alonso, Ekaterina N. Spirova, and David M. Holloway</i>	
Identifying Informative Genes for Prediction of Breast Cancer Subtypes	138
<i>Iman Rezaeian, Yifeng Li, Martin Crozier, Eran Andrechek, Alioune Ngom, Luis Rueda, and Lisa Porter</i>	
Predicting Therapeutic Targets with Integration of Heterogeneous Data Sources	149
<i>Yan-Fen Dai, Yin-Ying Wang, and Xing-Ming Zhao</i>	
Using Predictive Models to Engineer Biology: A Case Study in Codon Optimization	159
<i>Alexey A. Gritsenko, Marcel J.T. Reinders, and Dick de Ridder</i>	

Part IV: Protein: Structure, Function, and Interaction

Active Learning for Protein Function Prediction in Protein-Protein Interaction Networks	172
<i>Wei Xiong, Luyu Xie, Jihong Guan, and Shuigeng Zhou</i>	
Conditional Random Fields for Protein Function Prediction	184
<i>Thies Gehrman, Marco Loog, Marcel J.T. Reinders, and Dick de Ridder</i>	
Enhancing Protein Fold Prediction Accuracy Using Evolutionary and Structural Features	196
<i>Abdollah Dehzangi, Kuldip Paliwal, James Lyons, Alok Sharma, and Abdul Sattar</i>	

Exploring Potential Discriminatory Information Embedded in PSSM to Enhance Protein Structural Class Prediction Accuracy	208
<i>Abdollah Dehzangi, Kuldip Paliwal, James Lyons, Alok Sharma, and Abdul Sattar</i>	
Inferring the Association Network from p53 Sequence Alignment Using Granular Evaluations	220
<i>David K.Y. Chiu and Ramya Manjunath</i>	
Prediction of Non-genotoxic Hepatocarcinogenicity Using Chemical-Protein Interactions	231
<i>Chun-Wei Tung</i>	
 Part V: Motifs, Sites, and Sequences Analysis	
A Structure Based Algorithm for Improving Motifs Prediction	242
<i>Sudipta Pathak, Vamsi Krishna Kundeti, Martin R. Schiller, and Sanguthevar Rajasekaran</i>	
A Workflow for the Prediction of the Effects of Residue Substitution on Protein Stability	253
<i>Ruben Acuña, Zoé Lacroix, and Jacques Chomilier</i>	
Estimating Viral Haplotypes in a Population Using k-mer Counting	265
<i>Raunaq Malhotra, Shruthi Prabhakara, Mary Poss, and Raj Acharya</i>	
Fast Computation of Entropic Profiles for the Detection of Conservation in Genomes	277
<i>Matteo Comin and Morris Antonello</i>	
Author Index	289